REMARKS

Reconsideration and allowance in view of the foregoing amendment and the following remarks are respectfully requested.

Applicant notes that the Office Action lists claims 1-35 as pending in the application and simply comments to clarify that claims 22-25, 27-32 and 34-35 are pending in the application.

Rejection of Claims 22-25, 27, 29-32 and 34 Under 35 U.S.C. §103(a)

The Office Action rejects claims 22-25, 27, 29-32 and 34 under 35 U.S.C. §103(a) as being unpatentable over Ezzat et al. ("Visual Speech Synthesis by Morphing Visemes") ("Ezzat et al.") in view of Jiang et al. ("Visual Speech Analysis with Application to Mandarin Speech Training") ("Jiang et al.") in view of Hon et al. ("Automatic Generation of Synthesis Unites for Trainable Text-to-Speech Systems") ("Hon et al."). Applicant traverses this rejection and respectfully submits that first, one of skill in the art would not have sufficient motivation or suggestion to combine these references and secondly, even if combined, these references fail to teach each limitation of the claims.

To establish a *prima facie* case of obviousness, the Examiner must meet three criteria. First, there must be some motivation or suggestion, either in the references themselves, or in the knowledge generally available to one of ordinary skill in the art, to combine the references. Second, there must be a reasonable expectation of success, and finally, the prior art references must teach or suggest all the claim limitations. The Examiner bears the initial burden of providing some suggestion of the desirability of doing what the inventor has done. "To support the conclusion that the claimed invention is directed to obvious subject matter, either the references must expressly or impliedly suggest the claimed invention or the examiner must present a convincing line of reasoning as to why the artisan would have found the claimed invention to have been obvious in light of the teachings of the references." MPEP 2142.

If the proposed modification or combination of the prior art would change the principle of operation of the prior art invention being modified, then the teachings of the references are not sufficient to render the claims *prima facie* obvious. *In re Ratti*, 270 F.2d 810, 123 USPQ 349 (CCPA 1959). MPEP 2143.01.

Furthermore, if the examiner determines there is factual support for rejecting the claimed invention under 35 U.S.C. 103, the examiner must then consider any evidence supporting the patentability of the claimed invention, such as any evidence in the specification or any other evidence submitted by the applicant. The ultimate determination of patentability is based on the entire record, by a preponderance of evidence, with due consideration to the persuasiveness of any arguments and any secondary evidence. *In re Oetiker*, 977 F.2d 1443, 24 USPQ2d 1443 (Fed. Cir. 1992). The legal standard of "a preponderance of evidence" requires the evidence to be more convincing than the evidence which is offered in opposition to it. With regard to rejections under 35 U.S.C. 103, the examiner must provide evidence which as a whole shows that the legal determination sought to be proved (i.e., the reference teachings establish a *prima facie* case of obviousness) is more probable than not. MPEP 2142.

The test for obviousness is what the combined teachings of the references would have suggested to one of ordinary skill in the art, and all teachings in the prior art must be considered to the extent that they are in analogous arts. Where the teachings of two or more prior art references conflict, the examiner must weigh the power of each reference to suggest solutions to one of ordinary skill in the art, considering the degree to which one reference might accurately discredit another. *In re Young*, 927 F.2d 588, 18 USPQ2d 1089 (Fed. Cir. 1991). MPEP 2143.01.

The mere fact that references can be combined or modified does not render the resultant combination obvious unless the prior art also suggests the desirability of the combination. *In re Mills*, 916 F.2d 680, 16 USPQ2d 1430 (Fed. Cir. 1990).

Applicant respectfully submits that by a preponderance of the evidence one of skill in the art would not have sufficient motivation or suggestion to combine these references. Applicant will first discuss the teachings of Ezzat et al. Ezzat et al. teach a visual speech synthesis approach by morphing visemes. As noted in earlier arguments by Applicant, Ezzat et al. present an application called MikeTalk which is a text-to-audio visual speech synthesizer that converts input text into an audio visual speech stream. The goal of Ezzat et al. is to develop the text-to-audio visual speech synthesizer that is similar to a text-to-speech synthesizer except that it converts the text into an audio speech stream but also produces an accompanying visual stream which comprises a talking face that annunciates the text. The Office Action refers to Figure 1 of Ezzat et al. and the first column of page 46 Applicant notes that this reference does teach audio input and audio and video output.

The Office Action asserts on page 2 that Ezzat et al. teach the step of "selecting candidate image samples utilizing the target feature vector to generate a photorealistic animation of the object" at the top of the first column on page 51 and in the Abstract. Applicant notes that Ezzat et al. on page 51, first paragraph of the first column, teach a flow concatenation feature in which Ezzat et al. discuss the issue of the intermediate frames that lie between chosen viseme images. The pixel motions between these consecutive frames are small and hence the gradient-based optical flow method is able to estimate the displacements. Ezzat et al. teach computing a series of consecutive optical flow vectors between each intermediate image and its successor and concatenate them all into one large flow vector that defines the global transformation between the chosen visemes. Applicant notes that on page 3 of the Office Action the Examiner appears to assert that the visemes discussed in Ezzat et al. represent a generic facial image that can be used to describe a particular sound and the flowvectors contain "visual and sound features" that are used in conjunction with the viseme. Applicant traverses this analysis and notes that Section 5.2

of Ezzat et al. discusses optical flow and how the apparent motion is captured in a two-dimensional array of displacement vectors. The reference to "optical" flow vectors on pages 50 and 51 of Ezzat et al. have no reference to including sound features inasmuch as it appears that the optical flow vectors are utilized after the selection of viseme images. Accordingly, Applicant respectfully traverses the broadening interpretation of the optical flowvectors within Ezzat et al. to include sound features. The inclusion of sound features is not taught or suggested within Ezzat et al.

Next, the Office Action correctly notes that Ezzat et al. fail to teach the step of obtaining, for each frame in a plurality of end frames of an object animation, a target feature vector comprising visual features and non-visual features associated with the object animation. The Office Action asserts that Jiang et al. teach this step in the Abstract and in Section 4.2. Applicant traverses this analysis and notes that the method, for example, of claim 22 is a method for synthesis of photo realistic animation. In contrast, Jiang et al. teach key information that is extracted from a region of interest. Extracted feature vectors are further discussed in Section 4.2 in connection with Figure 4. In other words, Jiang et al. focus on extracting information from existing images. Jiang et al. simply do not teach synthesizing photorealistic animation. Rather, as has previously been noted, the key information that is extracted and the extracted feature vectors that are taught in Jiang et al. relate to visual, speech and analysis. Applicant highlights Figure 1 of the STODE system overview in which the student is illustrated with a camera with the region of interest detection module which passes information to the feature extraction module which then passes information to the lip model fitting module. Applicant recognizes in response to the Examiner's arguments on page 7 that Jiang et al. does mention visual feedback generation, but respectfully traverses the Examiner's response by noting that the issue of obviousness does not require one reference to "explicitly" teach away from the combination but rather the overall

suggestive power of a reference must be analyzed to determine whether by a preponderance of the evidence one of skill in the art would have sufficient motivation or suggestion to combine the references.

Applicant simply notes that the vast majority of Jiang et al. focuses on the visual speech analysis component and discusses the method of analyzing realtime lip tracking and feature extraction and multistate lip modeling using a time delay neuro-network the visual and audio images of a student. 95% of the reference focuses on the analysis component and a short paragraph at the end mentions that an artificial talking head that is cloned from the users own images may be animated in realtime and controlled from the visual speech analysis system. Jiang et al. also explicitly reference that generating corrected lip motions will rely on successful speech-to-lip movement synthesis techniques which "exceed the scope of this paper." Accordingly, on the balance, Applicant notes that it is clear that the visual feedback component of this paper is quite minor and explicitly beyond the scope of what is discussed in the paper. As a result, the overall suggestive power to one of skill in the art would certainly weigh against combining the techniques disclosed herein because the vast majority of this reference focuses on analysis. In sum, the Examiner appears to rely on the possibility of the combination because both references teach analysis and playback of audio video stream components. However, Applicant's analysis moves beyond what can be combined to what the suggestive power of each reference actually is and respectfully submits that by a preponderance of the evidence one of skill in the art would not have sufficient motivation to combine Jiang et al. with Ezzat et al.

Furthermore, on page 3 of the Office Action, the Examiner asserts that Jiang et al. teach that one advantage to obtaining feature vectors is to help children improve their speech pronunciation, citing Section 5, by providing audio-visual feedback. However, as eluded to above, Applicant respectfully submits that the overall purpose of Jiang et al. which is to provide

speech analysis for children to improve their speech pronunciation actually would urge one of skill in the art away from the teachings of Ezzat et al. inasmuch as the technical details found within Jiang et al. relate primarily to the analysis component rather than the visual speech synthesis component. Again, the Examiner's fundamental conclusion is that Jiang et al.'s advantage in obtaining feature vectors is in order to help children improve their speech pronunciation by providing audio-visual feedback does not adequately equate with the actual suggestion within Jiang et al. in which the paper explicitly notes that providing correct feedback and successful speech-to-lip movements which would be presented to the user "exceed the scope of this paper." Thus, it is clear that Jiang et al.'s teaches with regard to helping children improve their speech pronunciation by providing audio-visual feedback is beyond the scope of Jiang et al. and thus, one of skill in the art would recognize that important component of audio-visual feedback is not included within the teachings of Jiang et al. Thus, the express teachings here send one of skill in the art elsewhere to look for the audio-visual feedback that has the benefits identified by the Examiner on page 3 of the Office Action. For this additional reason, Applicant submits that one of skill in the art would not have sufficient motivation to combine Jiang et al. with Ezzat et al.

Next, the Office Action appropriately concedes that Ezzat et al. fail to teach the claimed step of an audio/video unit selection process in which the longest possible candidate image sample is selected. The Office Action asserts that Hon et al. teaches this claimed feature of using a multiple-instant system to construct long-units for frequent words and phrases that will achieve optimal concatenation quality. See page 296, column 1. Applicant respectfully traverses the analysis and notes that the Office Action does not even assert that Hon et al. teach the step of selecting a longest possible candidate image sample. At best, Hon et al. teach constructing long audio units for frequent words and phrases. Accordingly, even if these references were

combined, there is no teaching regarding a selection process in which a longest possible candidate image sample is selected. Applicant also questions the inclusion of a reference to Brand in this rejection. Brand is not listed as one of the prior art references used to reject claims 22-25, 27, 29-32 and 34 and thus, Applicant respectfully requests a clarification on whether Brand is being recited to reject these claims or not.

The Office Action concludes on page 4 that it would be obvious to one of skill in the art to combine Hon et al. with the combination of Ezzat et al. and Jiang et al. and that the advantage is that from the teachings of Hon et al. unit selection features selected from a database of a large amount of candidates can produce optimal concatenation quality. Applicant respectfully traverses this analysis and highlights that Hon et al. teach only unit selection in the context of text-to-speech. As noted in the title and the Abstract, Hon et al. focus on the whistler text-to-speech engine that was designed so that researchers can automatically construct model parameters from training data. In the teachings throughout this reference, it is clear that there is no teaching or suggestion that the speech unit selection process could be applicable or even possible in the context of photorealistic animation. In this regard, Applicant also submits that these are non-analogous arts inasmuch as the technologies involved in selecting samples for photorealist animation differ dramatically from selecting speech units for text-to-speech systems. One of skill in the art would certainly recognize these differences and not have sufficient motivation from the exclusive speech unit selection process taught in Hon et al. to incorporate that teaching into Jiang et al. and Ezzat et al.

Applicant further submits that even if the features of Hon et al. were incorporated into Ezzat et al. and Jiang et al., wherein the long units referenced in Hon et al. merely relate to words and phrases and thus, are exclusively speech oriented, that such combination would fail to teach each limitation of the claims. For example, if the teachings of Hon et al. were incorporated into

Jiang et al. and Ezzat et al. then the only modify component of a photorealistic animation of Jiang et al. and/or Ezzat et al. would be the speech component. Hon et al.'s teachings would provide long speech units for frequently used words and phrases, but would add nothing regarding images. However, what remains is the gap which is filled by the limitations of the claims in which the longest possible candidate image sample is selected. Accordingly, for this additional reason, Applicant submits that not only would one of skill in the art lack sufficient motivation or suggestion to combine Hon et al. with animation related references. Further, Applicant submits that even if these references were combined, they still fail to teach each claim limitation and the Office Action has not established the case where each limitation of the claims is taught.

Applicant respectfully submits that there are numerous issues with the analysis in the Office Action that fail to establish that claims 22-25, 27, 29-32 and 34 are not patentable. Accordingly, Applicant submits that these claims are patentable and in condition for allowance.

Rejection of Claims 28 and 35 Under 35 U.S.C. §103(a)

The Office Action rejects claims 28 and 35 under 35 U.S.C. §103(a) as being unpatentable over Ezzat et al. in view of Jiang et al. in view of Hon et al. and further in view of Brand ("Voice Puppetry") ("Brand"). Applicants respectfully traverse this analysis and submit that based on the analysis above, that one of skill in the art would not have sufficient motivation or suggestion to combine Ezzat et al. with Jiang et al. and further it is clear that one of skill in the art would not have motivation to combine Hon et al. with Ezzat et al. and/or Jiang et al.

Applicant notes that it is easy to establish that one of skill in the art would not have sufficient motivation to combine Brand with the other references and specifically with Hon et al. Brand focuses on a method for predicting a control signal from another related signal. They call their application "voice puppetry", which is the process of generating full facial animation from

expressive information in an audio track. Applicant notes that in this reference animation is produced by using audio to drive the model which induces the probability of distribution over the manifold of possible facial motions. Thus, Brand teaches receiving an audio signal and using analysis of that audio signal in order to drive a control signal for controlling animation. One of skill in the art would not combine Brand with Hon et al. because the audio signal that is utilized for producing a realistic whole face action is completely drawn from a user speaking or some other audio track. One of skill in the art would realize that the audio signal is easily obtained in this context and that the analysis in Brand requires a standard audio track. In fact, on page 23 before Section 4.1, they note that their process is much like speech recognition except that the units of interest are facial states rather than phonemes. Thus, one of skill in the art would certainly be unlikely to utilize the speech which is obtained from a text-to-speech synthesis system in the context of Brand because of the reduced quality of the audio signal that is obtained from a synthesized speech voice wherein it is clear that from the teachings of Brand that what is contemplated is using speech from a live or recorded audio signal from a user.

Accordingly, the fundamental principles of either Brand or Hon et al. would have to be abandoned if one of skill in the art were to blend their teachings. For example, if one of skill in the art was reviewing Brand and then faced with the teachings of Hon et al., rather than incorporating the teachings of Hon et al., Applicant submits that one skill in the art would simply dismiss or avoid the teachings of Hon et al. because of the lack of need for a synthesized voice. In contrast, the entire purpose of Brand is to provide voice puppetry for live or recorded human speech. The Office Action on page 6 asserts that the benefit of combining Brand with Hon et al., Ezzat et al. and Jiang et al. is that Brand teaches the advantage of using an optimal Viterbi sequence with a large number of sequence states to reproduce the size to the most optimal ones in order to remove poor animation quality, citing column 1 on page 25. However, this benefit

cannot overcome the challenges set forth above in which one of skill in the art would more likely than not avoid the text-to-speech system of Hon et al. altogether given the focus and purpose of Brand. Therefore, Applicant submits that claims 28 and 35 are patentable and in condition for allowance.

CONCLUSION

Having addressed all rejections and objections, Applicant respectfully submits that the subject application is in condition for allowance and a Notice to that effect is earnestly solicited. If necessary, the Commissioner for Patents is authorized to charge or credit the **Law Office of Thomas M. Isaacson, LLC, Account No. 50-2960** for any deficiency or overpayment.

Respectfully submitted,

Date: February 12, 2007

By: 

Thomas M. Isaacson

Attorney for Applicant
Reg. No. 44,166
Phone: 410-286-9405
Fax No.: 410-510-1433

Correspondence Address:
Thomas A. Restaino
Reg. No. 33,444
AT&T Corp.
Room 2A-207
One AT&T Way
Bedminster, NJ 07921